Against Understanding (as a Condition on Explanation)*

Bradford Skow

1 Understanding as a Condition on Explanation?

It is often said that there is an important relationship between explanation and understanding. Like so many ideas about explanation, this one makes an appearance in Carl Hempel's "Aspects of Scientific Explanation" (1965), the founding document of contemporary thinking about the topic. In the second section of that essay Hempel introduces the Deductive-Nomological Model of explanation, according to which (I'll risk the reminder) an explanation of a fact X is a sound argument for X that essentially contains a law-stating premise. Hempel then writes:

> [A] D-N explanation answers the question '*Why* did the explanandum-phenomenon occur?' by showing that the phenomenon resulted from certain particular circumstances, specified in $C_1, C_2, \ldots, C_k$, in accordance with the laws $L_1, L_2, \ldots, L_r$. By pointing this out, the argument shows that, given the particular circumstances and the laws in question, the occurrence of the phenomenon *was to be expected*; and it is in this sense that the explanation enables us to *understand why* the phenomenon occurred. (337; original emphasis)

What is going on in this passage? In particular, why did Hempel think it important to say the things he says in that last sentence? One natural interpretation has Hempel assuming, in the background, a necessary condition on explanation:

*To appear in *Making Sense of the World: New Essays on the Philosophy of Understanding*, Stephen Grimm, ed., Oxford University Press.

> **The Explanation-Understanding Condition:** Something E is an explanation of some fact F only if someone who possesses E understands F—at least in normal circumstances.

Then, this interpretation continues, what Hempel does explicitly in this passage is assert (it would be too much to call it an argument) that the DN model satisfies the necessary condition the Explanation-Understanding Condition states.

The Explanation-Understanding Condition, if true, looks like it could be a really useful tool for evaluating theories of explanation. One basic way to evaluate a theory of explanation is to "directly check it against your intuitions": find a body of fact that the theory counts as an explanation, and ask yourself, *is* it an explanation? Or, find an explanation, and then ask yourself, does the theory counts it as an explanation? If the answer to either question is no, that's a mark against the theory.[1] But it would be nice to have other ways to evaluate a theory, besides directly checking it against intuitions. The Explanation-Understanding Condition provides another way. Find a body of fact E that the theory counts as an explanation of F, and check if someone could possess E without understanding F. If that is possible, then the theory must be false.[2] In this way, it looks like the phenomenon of understanding can be used as an independent check on theories of explanation.

The Explanation-Understanding Condition has been, and continues to be, widely accepted. Michael Friedman, for example, in his 1974 paper "Explanation and Scientific Understanding," asks "What is it about . . . scientific explanations . . . that [give] us understanding of the world?" (5). A page later he asks again "what is the relation between phenomena in virtue of which one phenomenon can constitute an explanation of another, and what is it about this relation that gives understanding of the explained phenomenon?" (6). Friedman's questions presuppose the Explanation-Understanding Condition. I could go on with examples like

[1] I take it that this is the method philosophers *should* call "checking your intuitions," even though my statement of the method never uses the word "intuition" or any similar word. I realize this is controversial, but that controversy is not relevant here. (I am largely in agreement with Williamson (2008) and Cappelan (2012) about the role of so-called intuitions in philosophy.)

[2] Of course, this method involves using your "intuitions" about whether someone in some hypothetical circumstances understands F.

this for a long time. Here are two more: in 1984 Elliott Sober asserted that "Explanations afford understanding" (78). In 2003 James Woodward asserted that "It is a plausible constraint on what an explanation is that it must be something that provides understanding" (179). (I will have more to say about these authors' use of the Explanation-Understanding Condition later.)

I am "against" the Explanation-Understanding Condition. On one interpretation, the condition is false. On another interpretation, it can't, after all, be any use for evaluating theories of explanation.

A certain idea has shadowed philosophical debates about explanation like an assassin. That idea is that theories of explanation really are, or really should be, theories of answers to why-questions. I believe in this idea. I believe that all that philosophers of science should want out of a theory of explanation is a theory of answers to why-questions. And I hold that thinking about the phenomenon of understanding is no help at all in evaluating theories of answers to why-questions.

2 Explanations as Answers

No theory of explanation a philosopher has proposed has ever really been a theory of *all kinds* of explanation. Instead they have always been, at best, theories of explaining *why*. Suppose Smith asks Jones why the moon is waning, and in response Jones explains why the moon is waning by asserting some propositions (presumably about, among other things, the earth's and moon's orbits). The DN model, and every theory that has followed, can be read as aiming to say something about what it took for Jones to have succeeded in explaining (to Smith) why the moon is waning. But now suppose that Smith asks Jones what a gene is, and in response Jones explains what a gene is by asserting some propositions (presumably about inheritance, maybe also about chromosomes and DNA). There is no way to read either the DN model or any theory of explanation that has followed as aiming to say anything about what it took for Jones to have succeeded in explaining (to Smith) what a gene is.

I said in the last paragraph that every theory of explanation since the DN model can be read as a theory of explaining why. That's not entirely true. Bas Van

Fraassen's theory of explanation, in *The Scientific Image* (1980), can't. His theory takes the form of a theory of why-questions and their answers. That's not the same thing as a theory of explaining why. Assuming that the "canonical form" of an answer to the question why Q is something of the form "Q because R,"[3] a theory of answers to why-questions will fill in the schema

(S1) Q because R iff ...

A theory of explaining why, by contrast, will fill in the schema

(S2) Person P1 explained (to person P2) why Q iff ...

Since explaining is a speech act, to complete (S2) is to lay down the conditions required for someone to have successfully carried out this speech act. One does not have to lay down such conditions to complete (S1).

What I just said might suggest that a theory of explaining why must do "more" than a theory of answers to why-questions, in the sense that a theory of explaining why must contain a theory of answers to why-questions as a part. This suggestion is not true. A theory of explaining why will certainly *make use* of the notion of an answer to a why-question. But it can do this without containing, or entailing, any particular *theory* of answers to why-questions. In support of this, here is how I think the true theory of explaining why, the true completion of (S2), will start:

(S3) P1 explained to P2 why Q iff P1 told P2 the answer to the question why Q in the following way/manner: ....

Right there after "told P2" is the reference to the answer to the question why Q. But (S3), while it mentions answers to why-questions, doesn't contain a theory of such answers. Of course (S3) is not itself a theory; it trails off into ellipsis. But I don't think that the true completion of (S3) will contain a theory of such answers either. The stuff that comes after "way/manner" will (obviously) specify a way of telling

[3]Certainly not all answers have this form; the answer to the question why that plant is closing its stomata may be that it is doing this in order to conserve water. I happen to believe that every "in order to" answer—sometimes called teleological answers—is equivalent to a because-answer, but will not defend this claim here. (I defend it in chapter 6 of (Skow 2016).)

someone a proposition, but I can't think of any way of specifying a way of telling that involves stating a theory of answers to why-questions.

When Van Fraassen proposed his theory of explanation, which, again, was a theory of why-questions, he did not say that the criteria for judging his theory were different from those for judging earlier theories of explanation, like the DN model. But if his is a theory of why-questions and their answers, while earlier theories were theories of explaining why, shouldn't the criteria be different? The answer to this question would be yes—if the question's presupposition, that earlier theories were theories of explaining why, were true. The fact that Van Fraassen did not say that his theory should be judged on different criteria is some evidence that he, at least, thought that this presupposition was false. Anyway, it is my view that the presupposition is false. Although earlier theories can be read as theories of explaining why, I don't think they should be read that way. I don't think they're best interpreted that way. It's not only Van Fraassen's theory of explanation that is really a theory of answers to why-questions; that's what all of them really are.[4]

I don't take these observations to be terribly novel or even all that revolutionary. They've been around in some form or another from the beginning: Hempel himself says that explanations are answers to why-questions.[5] But even if the idea that philosophers of science should really be after a theory of answers to why-questions is in fact widely accepted, it is not really taken to heart. For philosophers

[4]There are some exceptions. For example, Achinstein is explicit that the theory he presents in (Achinstein 1983) is a theory of the speech act of explaining. I think this makes his theory a kind of theory that philosophy of science should not be interested in. I will defend this claim below.

[5]However, Hempel did not say that every answer to a why-question was an explanation; he isolated (or tried to isolate) a special class of why-questions that he called "explanation-seeking" (see Hempel 1965: 335). I argue against the existence of such a special class, and defend more generally the idea that theories of explanation really are, or at least should be read as, theories of answers to why-questions, in chapter 2 of (Skow 2016).

Some of the distinctions I've been drawing were first drawn by Bromberger, in the 1960's. He observed that one can explain things other than answers to why-questions, and emphasized that explaining an answer is different from telling someone the answer; see the papers collected in (Bromberger 1992).

rarely phrase their theories as theories of answers to why-questions. They phrase them as theories of explanation, phrasing them as completions of one of the following:

- Fact X explains fact Y iff ...
- Fact X is an explanation of fact Y iff ...

At one level this is just a matter of terminology: there is, I think, a well-established practice of using the noun "explanation," in some contexts at least, as a general term for answers to why-questions (only). In some contexts, "I was confused about the photoelectric effect; then Professor Smith gave me the explanation" means "I was confused about why the photoelectric effect happens; then professor Smith told me (the answer to the question) why the photoelectric effect happens." Still, over-use of the word "explanation" can seduce us into false claims and mislead us into accepting bad arguments. The Explanation-Understanding Condition is a good example of this.

3 Against the Explanation-Understanding Condition

Here is the Condition again:

> Something E is an explanation of some fact F only if someone who possesses E understands F—at least in normal circumstances.

Is this right? That depends on how it is interpreted. The first thing we need to do, to make it more precise, is to re-word it so that it does not speak of *facts* as the things that are explained, or understood. We do talk of explaining or understanding facts, but it is misleading to talk this way in this context. An explanation of the fact that the moon is waning is, in this context, just an explanation of *why* the moon is waning. Similarly, to understand the fact that the moon is waning is, in this context, to understand *why* the moon is waning. So let's make the "why"s that belong in the statement of the Explanation-Understanding Condition explicit (from now on I'm going to leave off the qualification about normal circumstances):

(EU) Something E is an explanation of why Q only if someone who possesses E understands why Q.

Now I want to distinguish two readings of (EU). One reading takes "explanation of why Q," as it appears in (EU), to mean nothing more than "answer to the question why Q." Another reading takes talk of an "explanation of why Q" to be talk of acts of performing the speech act of explaining.

Here is the first reading:

(EU1) A proposition P is the answer to the question why Q only if anyone who knows P understands why Q.

I should note that to get (EU1) from (EU) I did more than replace reference to an explanation with reference to an answer. There's another difficult bit of wording in (EU): it speaks of "possessing" an explanation. Once we eliminate "explanation," we are left with talk of someone "possessing" an answer to a why-question. But what is it to "possess" an answer to a why-question? The most natural thought is that possessing an answer is just knowing that answer.

Now we have one interpretation of the Explanation-Understanding Condition. Do we have in it an interesting constraint on theories of answers to why-questions? I don't think so, because I don't think (EU1) is even true. Since knowing why Q is the same thing as knowing the proposition that is the answer to the question why Q, (EU1) amounts to saying that knowing why Q is sufficient for understanding why Q. But this claim is false. Understanding is a greater achievement than knowledge, not a lesser one.

Consider, for example, Lester. Lester has never taken a chemistry class, or studied the subject on his own. But like all of us he has heard people use the words "acid" and "base," and words related to them, like "acidic." He has heard people say things like "Lemon juice is acidic," and "Baking soda is basic." He's had enough exposure to these words for them to be part of his vocabulary. For example, when he called his doctor recently with a stomach ache, and she asked him if he's had anything acidic to eat or drink recently, he understood her question, and replied that he had had some orange juice earlier in the evening. Now the other day Lester's niece was showing him the lab experiment she'd done in her chemistry class. She dipped a piece of litmus paper into a liquid, and it turned red. Lester was very curious about this result. "Why did the litmus paper turn red?" he asked her niece. "Because this stuff I dipped it into is an acid" she replied.

Here Lester's niece told him the correct answer to the why-question he asked. Moreover, she knew that that was the correct answer, having studied this experiment in her chemistry class. Furthermore, Lester has no reason to doubt her testimony. So after hearing his niece's answer Lester knew that the paper turned red because it was dipped it in acid. In other words, Lester knew why the litmus paper turned red. But, I claim, he did not *understand* why the litmus paper turned red. This is a counterexample to (EU1).

What more would it take for Lester to understand why the litmus paper turned red? I am not really sure. It doesn't really matter for what is to follow. But for what it is worth, here is some speculation. One possibility is that Lester needs more knowledge. He needs to know more than just the answer to the question why the litmus paper turned red in order to understand why the litmus paper turned red. Maybe what he needs to know, in addition, is some chemistry. More specifically, maybe he needs to know something about why dipping litmus paper into acids turns it red. If he knew that, he would know something about the "connection" between the event the why-question concerns (the color-change), and the fact offered in the answer to that why-question (that the paper was dipped in acid). And this is not just any old connection; instead, it appears to be the connection *in virtue of which* that answer counts as the answer. That is, the fact that acid causes litmus paper to turn red via such-and-such chemical process appears to be the (or at least an) answer to the question "Why is it true that the paper turned red because it was dipped in acid?" If these speculations are on the right track, then they suggest a generalization: maybe, in general, understanding why Q requires not just that one know the proposition that is the answer to the question why Q, but also know something about why that proposition is the answer. In terminology I introduce in (Skow 2016), the claim is that understanding why Q requires not just knowing some or all of the reasons why Q, but also knowing why those reasons are reasons.[6]

[6]Stephen Grimm holds that understanding why E happened requires knowing, not just that C (a cause of E) happened, but also knowing something about the "modal relationship" between C and E (Grimm 2014). It may be that the connection between C and E in virtue of which "because C happened" is the answer to the question why E happened just is the modal connection Grimm focuses on. But I will not pause here for a detailed comparison of the two views. Michael Strevens

So the first interpretation of (EU), namely (EU1), is not a principle that can be used to evaluate theories of answers to why-questions, because it is false. I turn now to a second interpretation of (EU). This interpretation focuses, not on the answer to the question why Q, but on the act of explaining why Q. It interprets (EU) as stating a necessary condition on having performed this speech act. Here is one way to state this condition:

(EU2) Person P1 explained to P2 why Q only if, as a result of what P1 did, P2 understood why Q.

What (EU2) is saying is that to count as having *explained* to P2 why Q, as opposed to, say, *telling* P2 why Q, one must do more than get him to believe the answer to the question why Q. What more that is (EU2) does not say—except that it is whatever it takes for P2 to end up understanding why Q.

So is (EU2) right? I honestly have no idea. I can see reasons to doubt it. There are plenty of principles that compete with it. Here are a few:

(EK) P1 explained to P2 why Q only if, as a result of what P1 did, P2 *knew* why Q.

(EP) P1 explained to P2 why Q only if, as a result of what P1 did, P2 *was in a position to know* why Q.

When I say that these principles compete with (EU2) I don't mean that they are incompatible with (EU2); in fact, assuming that understanding why Q entails knowing why Q, (EU2) entails each of them. They compete with (EU2) not for truth, but for acceptance. Maybe having one's audience (merely) know why Q is the strongest (relevant) necessary condition on successfully explaining why Q; maybe one's audience doesn't also need to have achieved so much as having come to understand why Q.

---

holds that understanding why E happened requires directly mentally apprehending the explanation of E, where knowing a proposition is not sufficient for directly mentally apprehending it (Strevens 2013). If we assume that "the explanation of E" here denotes the answer to the question why E happened, then I disagree; one need not directly mentally apprehend the answer, but one does need to know other propositions in addition to the answer.

When I compare (EU2) to (EK) and (EP), I just don't know how I would go about choosing between them.

But I also think it doesn't matter. What reason does a philosopher of science have to care whether (EU2), rather than one of the others, is true? The people who have reason to care are philosophers who aim to have a theory of the speech act of explaining. But, as I argued above, that's not what philosophers have been looking for under the heading of a theory of explanation.

Of course, it's possible that while philosophers of science *haven't* been looking for a theory of the speech act of explaining, they *should have* been.

Now if philosophers of science should be looking for a theory of explaining, that must be because a complete philosophy of science requires such a theory. But in fact I think that a complete philosophy of science shouldn't include a theory of explaining.

A complete philosophy of science would have to include a theory of explaining if, but only if, explaining were in some way part of the nature of science. Now we do sometimes say things that suggest that it is. Some philosophers say that one of the aims of science is to explain why things happen. They say this in opposition to those who say that science aims to provide (just) a comprehensive description of what happens. Whether they are right is contentious; this debate is part of the debate over "scientific realism." But just what claim is it that the debate is about? Does "science aims to explain why things happen" mean just "science aims to answer why-questions," or does it mean "science aims to inform the public (or whomever), by means of performing the speech act of explaining, why things happen"? Once this distinction is made, I think it's obvious that only the first interpretation has a chance of being true. Or I should say, it is only the first interpretation the scientific realists mean to affirm.

What exactly is scientific realism? It doesn't seem to be the same thing to everyone. But certain doctrines appear to be central: the doctrine that our current scientific theories are close to being true; the doctrine that it is a constitutive aim of science, as a practice, to produce true theories; and—*I* think—the doctrine that it is a constitutive aim of science to produce theories that are not just true, but also contain answers to why-questions. But I just can't see why anyone would think it

was a constitutive aim of science that its practitioners perform any particular speech acts.

Couldn't a mute Robinson Crusoe have been an excellent scientist? Maybe he finds himself alone in Geneva in a post-apocalyptic world. He fixes up the Large Hadron Collider and does a bunch of experiments. He improves on the current formulation of quantum field theory, devising a theory that is closer to the truth. He discovers the answer to the question why electrons, or Higgs bosons, do this or that. But he never says anything (not to anyone else, there being no one else there, and not to himself either). He never performs any speech acts. In particular, he never explains anything to anyone. Would any scientific realist really maintain that Crusoe was in some way defective as a scientist—that he was failing to pursue all of the aims of science?

Let's take stock. I distinguished two readings of the Explanation-Understanding Condition. One reading, (EU1), says that some fact about understanding is a necessary condition on something's being an answer to a why-question. But (EU1) is false. The other reading, (EU2), says that some fact about understanding is a necessary condition on someone's having (successfully) performed the speech act of explaining. Maybe (EU2) is true, but whether it is is of no interest to the philosophy of science.

4 An Example: Sober On Explanation and Causation

What I want to do next is look at a few examples of philosophers invoking the Explanation-Understanding Condition. We are now in a position to see the mistakes these invocations have led to.

I selected the examples to look at from important works in the philosophy of science, but the selection was otherwise random. There are a great deal more I could have looked at.

In his 1984 book *The Nature of Selection* Elliott Sober took up the question of whether an organism's fitness causes, or explains, its survival. This led him to a discussion of whether a disposition, say a sugar cube's being soluble, causes, or explains, its manifestation (the sugar cube's dissolving). Sober accepted that the

solubility causes the dissolving, but claimed that "we feel somewhat disappointed when told that the sugar lump dissolved in water because it was water-soluble" (77). Being told that the sugar lump is water-soluble is satisfying when we have asked what caused it to dissolve, but is (in his words) "disappointing" when we have asked why it dissolved. To reconcile the claim that the solubility is a cause with the claim that citing is does not make for a "satisfying" explanation, Sober set out to "disentangle issues of causation from issues of explanation":

> Consider the question "What caused Y?" One correct answer would be "the cause of Y." However, this answer might rightly be classified as unexplanatory, since it does not provide us with any better understanding of why Y occurred than we started with. Explanations afford understanding; therefore, claims asserting that X explains Y are true or false partly because of the relation of those two terms to a third—namely, *us*. Causality is not similarly "consciousness-dependent." Whether or not X caused Y is not in general influenced by whether *we* might find it interesting to be told this. (1984, 78; original italics)

Let's look closely at the argument Sober is offering. Here it is as he stated it:

1. "Explanations afford understanding."
2. "Therefore, claims asserting that X explains Y are true or false partly because of the relation of those two terms to a third—namely, *us*."
3. "Causality is not similarly 'consciousness-dependent.' "
4. [Implicit conclusion:] It is false that, or at least not always true that, if X is a cause of Y, then X explains Y.

I don't think that this is a very good argument. Its problems start with line 1. Line 1 is a version of the Explanation-Understanding Condition. I have proposed two readings of this condition. But the first reading, (EU1), is false, and the second, (EU2), is not an interesting thesis (from the perspective of the philosophy of science anyway). So the first premise of Sober's argument is either false or uninteresting.

The charge of uninterestingness might seem itself uninteresting. Surely what matters is how interesting Sober's *conclusion* is, not how interesting his premises are. So before dismissing the reading of Sober's argument that takes line 1 to be (EU2), we should at least see what the argument's conclusion looks like on that reading.

Since (EU2) comes from (EU) by interpreting talk of explanations as talk of the speech act of explaining, we should interpret line 4 to also make a claim about the speech act of explaining. Here is one stab at an interpretation of line 4 like that:[7]

4*. The following is false: for any term Z (name or description) denoting a cause of E, someone can explain to someone else why E happened by saying ⌜E happened because Z happened⌝.

I think that 4* may well be true, but that whatever interest it has does not essentially depend on its focus on the speech act of explaining. Here is how I would put the thought main Sober wants to get at in the quoted passage above, that 4* seeks to capture: if I ask you why some event E happened, and you reply that E happened because its cause happened (let's pretend that E has only one relevant cause), you will have failed in some way to respond to my question appropriately.

Now I can see a route one might try to take from this claim to an interesting conclusion, a conclusion about what can and cannot be an answer to a why-question. Here is how it goes, starting from the claim I ended the last paragraph with:

5. If, in response to the question why E happened, someone X responds by asserting that E happened because its cause happened, then X will have given a bad response to the question.

6. If, by asserting P, someone gives a bad response to a question, then P is not an answer to that question.

7. So the proposition that E happened because its cause happened is not an answer to the question why E happened.

[7] I've had to promote one schematic letter to a variable; I've also changed which letters I use for schematic letters/variables to make the conclusion easier to read.

8. So even though the cause of E is a cause of E, it is false that E happened because its cause happened; the schema "if Z is a cause of E, then E happened because Z happened" has false instances.

I want to make two points about this argument. The first is about who this argument is supposed to target. Elsewhere in his book Sober takes issue with the thesis that to answer the question why E happened one must cite, or describe, causes of E. The argument (5)–(8), if sound, would show this thesis to be false, *under one interpretation.* That interpretation says that ⌜E happened because R⌝ is true if and only if the sentence R in some sense describes the occurrence of a cause of E. But this does not seem to me a strong point against those who think that answers must cite or describe causes. They will just opt for a weaker interpretation of their thesis; they will find a principled way to restrict which descriptions of the occurrences of E's causes are allowed as values of R.

My second point is that the argument's premise 6 is false. Asserting P can be a bad response even if P is an answer. How? Here is one way. It could be that P, while an answer, is not the answer the asker is looking for. I might watch a window break after being hit by a rock,[8] but not know whether the window was fragile, or whether instead it was made out of super-glass that had just one point of weakness, the point where the rock was lucky enough to hit. I ask why the window broke. If you tell me that it broke because it was hit by a rock, that's a bad response. But the problem with it is not that what you said was *false.* "Because it was hit by a rock" is true, and is (therefore) an answer to the question why the window broke. The problem with your response is, instead, that it was not the answer I was looking for.

The point can be made without using a why-question as an example. For any question, there are many true answers to that question which it would be wrong to assert when asked that question. Suppose you and I are in some Boston greenspace or other, but I, being from out of town, don't know which. Are we on the Common, or in the Public Garden, or what? I asked you where we are. If you reply "We are in Boston," you give a true answer to the question "Where are we?" But it was wrong of you to assert that answer. It's not the answer I was looking for. Similarly,

[8] After the *window* is hit by a rock, that is.

suppose I ask you who came to the party, and you answer "the people who came to the party came to the party." You haven't said something false. You've given me an answer to my question. But it's an answer I already knew. It's not the answer I wanted.

This could be what is going on with the response Sober is interested in, the response to the question why E happened that one gives by saying "E happened because its cause happened." That's certainly a bad response; but it could be bad not because it's not an answer, but only because it's not the answer the asker wanted. If this is right, then Sober's observation that it would be bad to say "because its cause happened" in response to the question of why E happened does not show that there is any restriction on what descriptions of E's causes can appear in an answer to the question why E happened.

I have said that "We are in Boston" and "The people who came to the party came to the party" are correct answers to the relevant questions. This might be challenged. (I believe there are theories of questions and answers on which they are not answers.) But I don't think Sober is in a position to challenge them. He accepts, right at the beginning of the long quotation above, that "The cause of E" is a correct answer to "What caused E?".

To sum up: Sober, in the quoted passage, tried to use a connection between explanation and understanding to reach a conclusion about when descriptions of the causes of E can, and cannot, constitute explanations of the occurrence of E. I have argued that his argument was based on a false claim about the connection between explanation and understanding, and also sketched a way that someone who thinks that any description of the causes of E can appear in an answer to the question why E happened can resist his conclusion.

5 Another Example: Woodward Against the DN Model

We saw Hempel claim that the DN model satisfies the Explanation-Understanding Condition. In *Making Things Happen* James Woodward turns this around: he argues that the DN model is *false* on the ground that it *fails* to satisfy the Explanation-Understanding Condition. Woodward in his argument focuses on explanations that

do not appear to cite any laws, such as Michael Scriven's canonical example,

> INK BOTTLE: The carpet is stained because Jones knocked over the ink bottle.[9]

Hempel's view about INK BOTTLE, and other, similar, examples, was, roughly, that the sentence in INK BOTTLE is true, even though no law-statements follow the word "because," because that sentence, or someone who uses that sentence, in some way conveys information about the DN argument that is the "ideal" explanation of the carpet stain. Woodard calls the strategy behind this response "the hidden structure strategy." The "hidden structure" is the ideal explanation; it is hidden because someone could truly assert that the carpet is stained because Jones knocked over the ink bottle without knowing what laws appear in the ideal DN argument. The hidden structure strategy is to say that a "because" statement can be true by "conveying information" about a structure that is hidden from (not fully known to) the person asserting that statement, or his audience.

Woodward does not like the hidden structure strategy, and thinks that any theory of explanation that makes use of this strategy will fall afoul of the Explanation-Understanding Condition. Here is how Woodward puts his argument:

> There is yet another reason for rejecting the hidden structure strategy. This derives from a general point about the epistemology of explanation and the connection between explanation and understanding . . . It is a plausible constraint on what an explanation is that it must be something that provides understanding. To say that certain information is "part" of an explanation or contributes to its explanatory import is to say that this information contributes to the understanding provided by the explanation. This in turn imposes an epistemic constraint on what information can be part of an explanation and can contribute to its explanatory import: such information must be epistemically accessible to those who use the explanation. Put slightly differently, the idea is that the features of the explanation that endow it with explanatory import—that make it an explanation—must be features that can be

[9] It became canonical by being quoted at length and discussed by Hempel (1965: 360).

> known or grasped or recognized by those who use the explanation; if not, it isn't in virtue of possessing those features that the explanation produces understanding. On this way of looking at matters, there is something deeply puzzling about the suggestion that claims like [INK BOTTLE] explain or convey understanding in virtue of providing information about the existence of some underlying epistemically hidden structure, whether a DN argument or an ideal explanatory text. The mere obtaining of this structure, independently of anyone's awareness of its existence, cannot be what accounts for people's judgment that, for example, the impact of the knee on the desk is explanatorily relevant to the tipping over of the inkwell. If this line of thought is correct, it seems to follow that to the extent that information about the laws or structures that underlie singular-causal (or other sorts of causal) claims is epistemically hidden from those who use such explanations, it cannot be that this information contributes to the explanatory import of these explanations. (179-80)

While I agree with Woodward that the DN model is false, I don't think there is a good argument against it here in this passage.

Let's walk through Woodward's argument slowly. Woodward starts with a version of the Explanation-Understanding Condition:

1. "It is a plausible constraint on what an explanation is that it must be something that provides understanding."

He also asserts a version of this condition about "parts" of explanations:

2. "To say that certain information is 'part' of an explanation or contributes to its explanatory import is to say that this information contributes to the understanding provided by the explanation."

Woodward then says that claim 2 "imposes an epistemic constraint on what information can be part of an explanation and can contribute to its explanatory import." The epistemic constraint that he infers from claim 2 is this:

3. Information can be part of an explanation and can contribute to its explanatory import only if it is "epistemically accessible to those who use the explanation."

Presumably the argument proceeds from here by appealing to the implicit premise:

4. Plenty of people use, or have used, the explanation in INK BOTTLE, to whom no law of nature is epistemically accessible.

The argument's conclusion then follows from lines 3 and 4:

5. Therefore, no law of nature is part of the explanation in INK BOTTLE, and no law of nature contributes to its explanatory import.

How does all this look if we eliminate "explanation" in favor of "answer to a why-question"? It is not entirely obvious how to translate each of the premises. Here is one attempt. Line 1 becomes (EU1):

1*. A proposition A is an answer to the question why Q only if someone who knows A understands why Q.

Line 2 takes line 1 and turns it into a claim about parts of an explanation, so line 2* should take line 1* and turn it into a claim about parts of an answer:

2*. If P is part of an answer A to the question why Q, then anyone who knows A−P ("A minus P," the information left over when P is "subtracted" from A) understands why Q to a lesser degree than someone who knows (all of) A.

Line 3 required parts of explanations to be epistemically accessible, and was inferred from line 2, so from line 2* we should infer a claim about parts of *answers* being epistemically accessible:

3*. If P is part of an answer A to the question why Q, anyone who answers the question why Q by asserting A knows P.

Let me pause briefly: premise 3 spoke both of being part of an explanation, and also of contributing to that explanation's explanatory import. But it's hard for me to

see why adding the stuff about explanatory import is needed; how could something be part of an explanation without contributing to that explanation's explanatory import? In terms of answers, how could something be part of an answer without contributing to that answer? As far as I can tell, it couldn't. So I'm leaving out the stuff about explanatory import in the *-ed interpretation of Woodward's argument.

Let's continue. There's one premise left, and then the conclusion. I'll put the whole argument here, with the last premise and the conclusion at the end:

1*. A proposition A is an answer to the question why Q only if someone who knows A understands why Q.

2*. If P is part of an answer A to the question why Q, then anyone who knows A−P understands why Q to a lesser degree than someone who knows (all of) A.

3*. (So,) if P is part of an answer A to the question why Q, anyone who answers the question why Q by asserting A knows P.

4*. Plenty of people who have answered the question why the carpet is stained by asserting the proposition Ink bottle did not know any law of nature.

5*. Therefore, no law of nature is part of the answer to the question why the carpet is stained in Ink bottle.

Is Woodward's argument, interpreted this way, any good? Premise 1*, again, is (EU1), which I have rejected. That's enough for the argument to fail. But let's look at the rest of it. Maybe there's a way to fix the argument, so that it doesn't rely on 1*?

I think there is. The argument's intermediate conclusion, line 3*, looks plausible to me. Surely if you assert A you know A—or at least this is usually the case. And if you know A you know (or at least are in a position to know) its parts.

Of course, the idea of a "part" of a proposition is not an ordinary one. It is some kind of technical notion.[10] But the paradigm case must be conjunctions: the proposition that X is part of the proposition that X and Y. In this case a proposition

[10]It receives a detailed development in (Yablo 2014).

entails each of its parts (though not vice versa). Presumably this holds in general: if P is part of a proposition A, then A entails P. And it is not too wild to assume that if you know A, and A entails P, you are at least in a position to know P. (Maybe this looks better if we require you to *know* that A entails P; but such niceties don't matter here.)

So, interestingly, the *-ed version of the argument arrives at a plausible intermediate conclusion, line 3*, an intermediate conclusion that, as I have stated it, says nothing about understanding; this despite the fact that it is reached from suspect claims about the connection between understanding why Q and being an answer to the question why Q.

In fact, the rest of the *-ed interpretation of Woodward's argument goes fine as well. I think that 4* is true: plenty of people *have* used INK BOTTLE without knowing any laws (or, at least, we can easily consider a hypothetical scenario in which this is so). And the conclusion 5* follows from 3* and 4*.

Despite all this, there is still a big problem with the *-ed argument. The problem is not that it is unsound; the problem is that its conclusion is not the conclusion Woodward wants. Sophisticated defenders of the DN model, or theories that descend from the DN model, Peter Railton for example (whose views Woodward discusses at length in the chapter from which I am quoting), do not say that some law of nature is *part of the answer* in INK BOTTLE. They say instead that INK BOTTLE is *itself* only part of a larger and more complete answer, and that that larger and more complete answer contains laws of nature.[11]

Is there a better interpretation of Woodward's argument, an interpretation that gets him to the right conclusion? Well there is at least a different interpretation; let us see if it is better.

If you look back at the long quotation above, you will see that after Woodward drew the intermediate conclusion 3 he wrote

> Put slightly differently, the idea is that the features of the explanation that endow it with explanatory import—that make it an explanation—must be features that can be known or grasped or recognized by those

[11] See, for example, (Railton 1981).

who use the explanation; if not, it isn't in virtue of possessing those features that the explanation produces understanding.

I actually do not think that this is conclusion 3 "put slightly differently." I think that this is a very different idea. To ask *what features of an explanation endow it with explanatory import*, or to ask *what features make it an explanation*, sounds, to me, like asking *why it is an explanation.* If we focus, as we should, on answers to why-questions, not explanation, the question we are attending to becomes the question of *why a given answer to a why-question is an answer.* What do Woodward's remarks suggest about the answer to this question? Here's what I think they suggest:

6. If P is the answer to the question of *why is it that R is the answer to the question why Q*, then anyone who asserts R as an answer to the question why Q must be in a position to know P.

Here P, the answer to the question of why R is the answer to the question why Q, describes the "features of the explanation"—that is, the answer—"that make it an explanation," or answer.

Unlike claim 3*, claim 6 *can* get us to the conclusion Woodward wants—if it is true. Let's first see how it gets Woodward where he wants to go, and then turn to the question of whether it is true.

The argument from 6 against the DN model is, I think, straightforward. We need just two more premises:

7. If the DN model is correct, then some law of nature is (at least part of) the answer to the question of why Ink bottle is the answer to the question why the carpet is stained.

8. Plenty of people who have answered the question why the carpet is stained by asserting the proposition Ink bottle did not know (and were not in a position to know) any law of nature. (This is premise 4* from the *-ed argument, slightly revised.)

From 6, 7, and 8, the desired conclusion follows:

9. The DN model is not correct.

What about premises 7 and 8? Line 8 is certainly plausible, for the same reason that premise 4* from the *-ed argument was. I myself don't know of any laws of nature that could be added to INK BOTTLE to turn it into a DN argument. The plausibility of line 7 is harder to assess. Rather than dwelling on line 7, though, I will skip to the end: ultimately this argument does not succeed, because its crucial premise, 6, is false.

At least that's what I think. I think that someone could know the answer to the question why Q, and could tell someone who wants to know why Q that answer, without having any idea *why* that answer is the answer.

Suppose Jones is on a tour of the Old Mill. The tour guide points to some charred wood, and says that these are the parts of the mill that burned in the fire of '03. Jones asks why the mill caught on fire in '03. The tour guide replies that it caught on fire because it was struck by lightning. Jones then persists in his questioning. He says, "that's great, but I have another question. Why is that the answer? Why is it that 'because it was struck by lightning' is the answer to the question of why the mill caught on fire?" The tour guide is baffled, and a little annoyed. "I don't know" he says. He gives Jones a look to get him to shut up. Still, despite not knowing the answer to Jones' second question, the tour guide knew *why the mill caught on fire.* He also successfully told Jones why the mill caught on fire.

It is important to distinguish the baffling question Jones asked from another one. Jones did not ask *why the answer is true.* If, after the tour guide had said "because it was struck by lightning," Jones had said, as my two-year-old automatically says in response to pretty much everything I say, "Why?", we would most naturally take him to be asking *why the mill was struck by lightning.* To ask that is to ask why the tour guide's answer is true. That's not the same as asking why it is an answer.

I extracted premise 6 from some things Woodward said. But he didn't just say some things that suggested 6; he also offered a little argument for those things. Maybe 6 will seem more plausible if we look at that argument? Here, again, is what Woodward wrote; I've italicized the relevant bit:

> the idea is that the features of the explanation that endow it with explanatory import—that make it an explanation—must be features that can be known or grasped or recognized by those who use the expla-

nation; *if not, it isn't in virtue of possessing those features that the explanation produces understanding.*

This argument is worthy of our attention not just because Woodward offers it, but also because it appeals to a connection between explanation and understanding, and such connections are the main concern of this paper. So what is the argument? I think it is a good idea to start by making the argument more explicit using Woodward's own terminology, before trying to see how to put it as an argument for 6. Here is what the argument seems to me to be:

10. The features of an explanation that make it an explanation are (also) features in virtue of which that explanation produces understanding.

11. The features of an explanation in virtue of which it produces understanding are features that those who use that explanation can know it to have.

12. Therefore, the features of an explanation that make it an explanation are features that those who use that explanation can know it to have.

The conclusion 12, of course, corresponds to premise 6 of the earlier argument:

6. If P is the answer to the question of *why is it that R is the answer to the question why Q*, then anyone who asserts R as an answer to the question why Q must be in a position to know P.

What we need to do is "translate" lines 10 and 11 so that they can serve as premises in an argument for 6. I don't think this is hard to do, in light of my earlier discussions of what could be meant by "explanation produces understanding," and my earlier claim that "X in virtue of W" can be rendered as "W is the answer to the question why X." I'll label the translations of lines 10 and 11 with *s:

10*. If P is the answer to the question of why is it that R is the answer to the question why Q, then P is the answer to the question of why anyone who knows R understands why Q.

11*. If P is the answer to the question of why anyone who knows R understands why Q, then anyone who offers R as an answer to the question why Q is in a position to know P.

6. (Therefore,) if P is the answer to the question of why is it that R is the answer to the question why Q, then anyone who asserts R as an answer to the question why Q must be in a position to know P.

This argument is valid, but the premises 10* and 11* don't seem to me to have anything going for them. Start with 10*: the problem with it is that it presupposes that it is *true* that anyone who knows the answer to the question why Q understands why Q. I've said why I think this is false. I also don't see much reason to accept premise 11*. I can see no reason to think that you can only answer a why-question if you know why the answer you're giving produces understanding in those who hear it.

I have been examining Woodard's argument that the DN model fails to respect some connection between explanation and understanding. I found two different arguments that Woodward might have been giving; I have argued that both of them fail.

6 Conclusion

The idea that there is an important connection between explanation and understanding is a common one. The alleged connection, vaguely stated, is that explanations produce understanding. If it existed, this connection would open up a new avenue for searching for the true theory of explanation: just investigate what things produce understanding. But I think that the promise that comes with this alleged connection is not fulfilled. On one precisification of the vague idea, the connection does not exist—knowing the answer to the question of why some event E happened is not sufficient for understanding why it happened. On another precisification the connection is this: someone has performed the speech act of explaining with respect to the question of why E happened only if their audience came to understand why it happened. Maybe so, but a theory of this speech act is not what philosophers of science have been after under the heading "theory of explanation."

After all this negativity I want to draw attention to the fact that my title is not "Against Understanding." While I have opposed a certain use of the notion of understanding in theorizing about answers to why-questions, I do not think that

it is a useless notion for other purposes, or that the question of what it takes to understand why some event E happened is not itself independently philosophically interesting. In fact at least one place where the notion of understanding may be important is not hard to find. I said earlier that one of the aims of science is to answer why-questions (at least, according to some scientific realists). But it is plausible that science aims at more. Certainly science aims not just to "have" the answer to the question of, for example, why the dinosaurs went extinct, in some thin sense of "have," where believing it on very weak evidence is enough. Science aims to *know* why the dinosaurs went extinct. But even this might undersell how lofty the aims of science are. I find it plausible that science aims to *understand* why the dinosaurs went extinct; more generally, it aims, for all the facts F in its domain, to understand why F obtains. Even if we do not need to understand understanding in order to figure out what it takes to be an answer to a why-question, we do need to if we want to understand science.[12]

References

Achinstein, Peter (1983). *The Nature of Explanation.* Oxford University Press.

Bromberger, Sylvain (1992). *On What We Know We Don't Know.* The University of Chicago Press and CSLI.

Cappelen, Herman (2012). *Philosophy Without Intuitions.* Oxford University Press.

Friedman, Michael (1974). "Explanation and Scientific Understanding." *The Journal of Philosophy* 71: 5-19.

Grimm, Stephen. "Understanding as Knowledge of Causes." in A. Fairweather (ed.), *Virtue Epistemology Naturalized: Bridges Between Virtue Epistemology and Philosophy of Science.* Springer: 329-345.

Hempel, Carl (1965). "Aspects of Scientific Explanation." In *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science.* The Free Press.

Railton, Peter (1981). "Probability, Explanation, and Information." *Synthese* 48: 233-256.

Sober, Elliott (1984). *The Nature of Selection.* University of Chicago Press.

[12]I'd like to thank Fred Feldman for discussing this material with me.

Skow, Bradford (2016). *Reasons Why.* Oxford University Press.

Strevens, Michael (2013). "No Understanding Without Explanation." *Studies in History and Philosophy of Science* 44: 510-515.

Van Fraassen, Bas C. (1980). *The Scientific Image.* Oxford University Press.

Williamson, Timothy (2008). *The Philosophy of Philosophy.* Wiley-Blackwell.

Woodward, James (2003). *Making Things Happen.* Oxford University Press.

Yablo, Stephen (2014). *Aboutness.* Princeton University Press.